Uncertainty Estimation and Decomposition based on Monte Carlo and Multimodel Photochemical Simulations

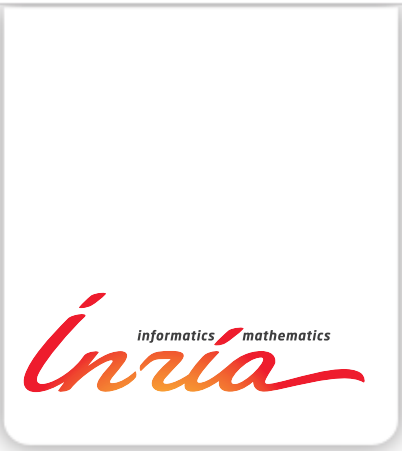
Damien Garaud, Vivien Mallet

Uncertainty Estimation and Decomposition based on Monte Carlo and Multimodel Photochemical Simulations

Damien Garaud , Vivien Mallet

RESEARCH REPORT
N° 7903
March 2012
Project-Team Clime

ISSN 0249-6399 ISRN INRIA/RR--7903--FR+ENG



Uncertainty Estimation and Decomposition based on Monte Carlo and Multimodel Photochemical Simulations

Damien Garaud * †, Vivien Mallet † *

Project-Team Clime

Research Report n° 7903 — March 2012 — 30 pages

Abstract: This paper investigates (1) the main sources of uncertainties in ground-level ozone simulations, (2) the best method to estimate them, and (3) the decomposition of the errors in measurement, representativeness and modeling errors. It first compares the Monte Carlo approach, solely based on perturbations in the input fields and parameters, with the multimodel approach, which relies on an ensemble of models with different chemical, physical and numerical formulations. Two ensembles of 100 members are generated for the full year 2001 over Europe. Their uncertainty estimations for ground-level ozone are compared. For both ensembles, we select a sub-ensemble that minimizes the variance of the rank histogram, so that it is supposed to better represent the uncertainties. The multimodel (sub-)ensemble shows more variability and seems to better represent the uncertainties (especially for the localization of the covariances) than the Monte Carlo (sub-)ensemble. The main sources of the uncertainties originating in the input fields and parameters are then identified with a linear regression of the output ozone concentrations on the applied perturbations. The uncertainty ranges due to the different input fields and parameters are computed at urban, rural and background observation stations. For both the multimodel ensemble and the Monte Carlo ensemble, ozone boundary conditions play an important role, even at continental scale; but many other fields or parameters appear to be a significant source of uncertainty. The discrepancies between observations and model simulations are due to measurement errors, representativeness errors and modeling errors (i.e., shortcomings in the model formulation or in its input data). Using two independent methods, we estimate the variance of the representativeness errors. We conclude that the measurement errors are comparatively low, and that the representativeness errors can explain at least a third of the variance of the discrepancies.

Key-words:

* CEREA, joint laboratory cole des Ponts ParisTech - EDF R&D, Universit Paris-Est
† INRIA

RESEARCH CENTRE
PARIS – ROCQUENCOURT

Domaine de Voluceau, - Rocquencourt
B.P. 105 - 78153 Le Chesnay Cedex

Contents

1 Introduction

Many studies and decision making are based on modern chemistry-transport models that solve 3D reactive transport equations for the main atmospheric chemical species. The primary output of these models is 3D concentrations of the chemical species against time. A next step in the development of air quality modeling systems should be to provide, along with the concentrations, an estimate of their uncertainty. The uncertainty can be significantly high, to the point of changing or at least softening conclusions drawn from the concentrations alone.

In order to estimate the uncertainty on a model's output, all the sources of this uncertainty should a priori be taken into account and propagated through the reactive transport equations. The ideal target would be the probability density function of the chemical concentrations, and its time evolution. Computing this probability density function is an impossible task, considering a model's state vector with easily one million or ten million components. In practice, the most accurate uncertainty estimations are derived from ensemble simulations. There are two main approaches: Monte Carlo simulations and multimodel simulations. In the former, the input data and parameters of the model are randomly perturbed in each member (i.e., simulation) of a typically 50- or 100-member ensemble [e.g., Hanna et al., 1998, 2001; Beekmann and Derognat, 2003; Boynard et al., 2011]. It requires to have knowledge on the probability distributions of the input data and parameters, and to have enough computational power to reasonably sample these distributions. The other approach relies on different chemistry transport models that are based on different physical, chemical and numerical formulations [e.g., Delle Monache and Stull, 2003; Mallet and Sportisse, 2006; McKeen et al., 2007]. This approach can be easily combined with Monte Carlo simulations since the input data of the members of a multimodel ensemble can be randomly perturbed. The combined approach can take into account all sources of uncertainties, from the model formulation itself (what Pinder et al. [2009] call structural uncertainty) and from the input data.

The Monte Carlo approach is rather easy to implement since only the input data and parameters are modified, within a single model. The multimodel approach has a higher implementation cost since it involves models with varying requirements: different input data, a range of resolutions, different chemical species, ... However this approach, when combined with perturbations in the input data, provides richer ensembles. One may wonder whether the better quality of the multimodel ensembles is worth the additional implementation effort. This raises the question of the main sources of uncertainties and of the merits of each approach with respect to the representation of the uncertainty sources. In this paper, we try to bring some answers by comparing a Monte Carlo ensemble with a multimodel ensemble. In section 2, we describe and compare the generation of two 100-member ensembles, one with the Monte Carlo approach and another with the multimodel approach, for the full year 2001. Uncertainty estimations computed with the empirical standard deviations are then analyzed. In order to improve these estimations, we also carry out an a posteriori calibration, using the observations and the method introduced in Garaud and Mallet [2011].

In section 3, we identify which input data and parameters are the main sources of the uncertainties in the output concentrations. We carry out the

identification for both Monte Carlo and multimodel ensembles. For the main sources of uncertainties, the impact on the output concentrations is estimated. The results point out which input fields or parameters should receive more attention in the generation of an ensemble.

In section 4, the discrepancy between observations and models' concentrations are decomposed into three errors: (1) measurement error due to instrument limitations, (2) modeling error originating from the shortcomings and uncertainties in the simulations, and (3) the representativeness error due to the misrepresentation of point observations by model's concentrations that are averaged over a grid cell. The latter is seldom estimated and yet proves to be quite significant. It is consistently estimated by two independent methods, one solely based on observations, and another based on the observations and any model from the ensembles.

2 Comparison of Monte Carlo and Multimodel Ensembles

In this section, we compare the uncertainty estimations that can be derived from Monte Carlo simulations and a multimodel approach. After each generation of an ensemble, a calibration is carried out, in order to select a sub-ensemble that better estimates the uncertainty than the full ensemble. This lets us to produce more accurate uncertainty estimations and to evaluate the estimation quality from both approaches.

2.1 Generation of the Ensembles

Air quality models produce estimations of the pollutant concentrations that can be highly uncertain because of (1) shortcoming in their chemical and physical formulations (turbulence modeling, deposition velocities, chemical mechanism, ...), (2) uncertainties in the various input data (meteorological fields, emission sources, boundary conditions, ...), and (3) the numerical approximations (numerical schemes, time step, vertical resolution). Considering all these uncertainty sources, many different models and associated simulations can be carried out with a great variety of results. In order to explore these possible results, two classical approaches are the Monte Carlo simulations that solely rely on perturbations in the uncertain input data or model parameters, and the multimodel ensembles that try to take into account all uncertainty sources.

In the multimodel ensemble approach, each simulation is built with a given combination of chemical/physical parameterizations, numerical schemes and input data. This ensemble can be derived from models developed in different teams or inside a flexible modeling platform. In this paper, we rely on the latter, with the air quality modeling system Polyphemus [Mallet et al., 2007] and the automatic generation of large ensembles described in Garaud and Mallet [2010]. In fact we use the same 101-member ensemble as described in Garaud and Mallet [2010] for photochemistry simulation over Europe during 2001. Every member of the ensemble has a unique combination of (1) perturbed input fields such as wind field, temperature or boundary conditions, (2) physical parameterizations (e.g., for the vertical diffusion coefficient or the chemical mechanism) and (3) numerical approximations such as time step, vertical resolution or the

first layer height, 40 m or 50 m. A total of thirty alternatives are available for the generation of a single model. Each member of the ensemble is defined by a random selection of one option per alternative.

In this paper, the Monte Carlo ensemble generation is inspired by Hanna et al. [1998, 2001] and Beekmann and Derognat [2003]. The uncertain input data is perturbed independently for each member of the ensemble, but each member relies on the same model. This model is referred to as the reference model in Garaud and Mallet [2010]. It uses common options of the Polyphemus system for photochemical simulations: RACM chemical mechanism, Troen&Mahrt parameterization to compute vertical diffusion coefficients, deposition velocities computed with the Zhang parameterization [Zhang et al., 2003], a vertical resolution with 5 levels from 50 m to 3000 m, a time step equal to 600 s. The perturbations are randomly sampled assuming that the input fields and parameters are normally or log-normally distributed with a given standard deviation. Table 1 lists the perturbed variables together with their distributions (normal or log-normal) and the associated uncertainties. Assume that p is a scalar to be perturbed. p is either a parameter or the point value of a field at a given time and location. In the case of a log-normal distribution, the perturbed value will be $\tilde{p} = p\sqrt{\alpha}^{\gamma}$ where α is given in the column "Uncertainty" of table 1 and γ is sampled according to the standard normal distribution. For a normally-distributed field or parameter, the perturbed value will be $\tilde{p} = p + \frac{1}{2}\alpha\gamma$. For temperature, a relative uncertainty is provided, so that the perturbed value will be $\tilde{p} = p(1 + \frac{1}{2}\alpha\gamma)$. Note that, for a given field, the same perturbation (i.e., $\sqrt{\alpha}^{\gamma}$, $\frac{1}{2}\alpha\gamma$ or $1 + \frac{1}{2}\alpha\gamma$) is applied for all times and locations; γ is therefore generated by a pseudo-random number generator just once for each field or parameter (and for each member). The uncertainties for the reaction rates of RACM are adapted from Beekmann and Derognat [2003] who use MELCHIOR chemical mechanism [Lattuati, 1997]. A comparison between the MELCHIOR and RACM reactions was carried out in order to assume the same uncertainty for almost all corresponding reaction rates. In total, we carried out the Monte Carlo runs to generate a 100-member ensemble.

A multimodel ensemble has a more detailed representation of the uncertainties as it can include perturbations in the input data (just like Monte Carlo simulations), but also takes into account the uncertainties due to the physical, chemical and numerical formulations. For instance, while the Monte Carlo simulations will take into account the shortcomings in the modeling of the turbulence with perturbations in the vertical diffusion coefficients, a multimodel approach will introduce several parameterizations for the vertical diffusion, each of which is physically consistent with the meteorological fields. A change of parameterization leads to much more relevant differences (e.g., increased vertical diffusion only in unstable conditions) than a simple multiplication of a reference diffusion field. In our case, the perturbations in the multimodel ensemble are mostly applied to the same fields and parameters as in the Monte Carlo simulations, except for the reaction rates that are perturbed only in the Monte Carlo simulations.

Table 1: Input data uncertainties for Monte Carlo simulations. The column "Uncertainty" reports the interval with 95% confidence whose median coincides with the unperturbed data. If the uncertainty is α and the initial parameter is p, every sample will lie in $[p/\alpha, \alpha p]$ for a log-normal distribution and $[p-\alpha, p+\alpha]$ for a normal distribution. For temperature, the interval is $[(1-\alpha)p, (1+\alpha)p]$.

Field	**Uncertainty**	**Distribution**
Meteorological Fields		
Wind speed	1.5	Log-normal
Wind angle	20°	Normal
Temperature	1%	Normal
Vertical diffusion coefficient	1.9	Log-normal
Attenuation	1.3	Log-normal
Emissions		
Anthropogenic NO_x	1.7	Log-normal
Anthropogenic VOCs	1.7	Log-normal
Biogenic VOCs	2.	Log-normal
Chemical rates – RACM [Stockwell et al., 1997]		
Reaction rates 128, 130, 132, 133		
136, 137, 143, 146, 147		
151, 152, 155, 156, 162		
166, 167, 168, 232, 236	1.3	Log-normal
Reaction rates 127, 129	1.2	Log-normal
Others reaction rates	1.1	Log-normal
Others		
Photolysis rates	1.4	Log-normal
Boundary conditions NO_x	3.	Log-normal
Boundary conditions	2.	Log-normal
Deposition velocities	1.7	Log-normal

2.2 Comparison of the Non-Calibrated Ensembles

The results of the two ensembles are studied for ozone peaks. The evaluation is carried out using the observation network Airbase[1]. The database, managed by the European Environment Agency, provides ground-level ozone observations at 210 rural background stations, 702 rural stations, 647 suburban stations and 1324 urban stations over Europe. In order to select stations which are representative of the ozone peak concentration at the model scale (half a degree in the horizontal), only rural and background stations are kept in this section. In order to select the most reliable stations, we exclude the stations that fail to provide observations for over 10% of all dates (i.e., all days in the year 2001). Following usual recommendations [Russell and Dennis, 2000; Hogrefe et al., 2001; US EPA, 1991], a cut-off is applied to the observations. Observations below 40 $\mu g\,m^{-3}$ are discarded. In total, we keep about 1.1 10^5 observations for ozone peaks during the year 2001.

Figure 1 shows the models' performances on a single figure, using a Taylor diagram for the two ensembles. A Taylor diagram [Taylor, 2001] takes into account the standard deviation of observations and the correlation between each simulation and the observations. The radial coordinate is the simulation standard deviation normalized by the standard deviation of the observations. The azimuthal coordinate corresponds to the arcosine of the correlation between each simulation and the observations. The performances of the simulations from the multimodel ensemble show higher spread than the performances of the simulations from the Monte Carlo ensemble. In the Monte Carlo case, all the simulations have a standard deviation less than the standard deviation of the observations. Moreover, almost all correlations are between 0.6 and 0.7. On the contrary, the range of correlations in the multimodel ensemble is wide, and several simulations show a higher variability than the observations.

The mean of the ensemble standard deviations (at observation stations) is 22.3 $\mu g\,m^{-3}$ for the multimodel ensemble and 19.9 $\mu g\,m^{-3}$ for the Monte Carlo ensemble. It shows that the simulations from the Monte Carlo experiment have a lower variability than those of the multimodel ensemble. It is noteworthy that the "best" model (in terms of RMSE and correlation) from the multimodel ensemble better compares to the observations than the "best" model from the Monte Carlo ensemble. The "best" simulation in the multimodel ensemble has a RMSE equal to 20.5 $\mu g\,m^{-3}$ and a 0.735 correlation whereas the "best" Monte Carlo simulation has a 21.8 $\mu g\,m^{-3}$ RMSE and a 0.679 correlation.

2.3 Comparison of the Calibrated Ensembles

The two ensembles take into account many sources of uncertainties, but the actual estimation of uncertainty that can be derived from them may not be reliable. In order to assess the quality of the uncertainty estimation by the ensembles, we compute the rank histogram [Anderson, 1996; Hamill and Colucci, 1997; Talagrand et al., 1999]. Each observation is given a rank which is the number of members that simulate a concentration lower than the observation. The rank is zero if the observation is below the lower envelope of the ensemble, and it is equal to the total number of members if the observation is above the

[1] http://air-climate.eionet.europa.eu/databases/airbase/airbasexml/index_html

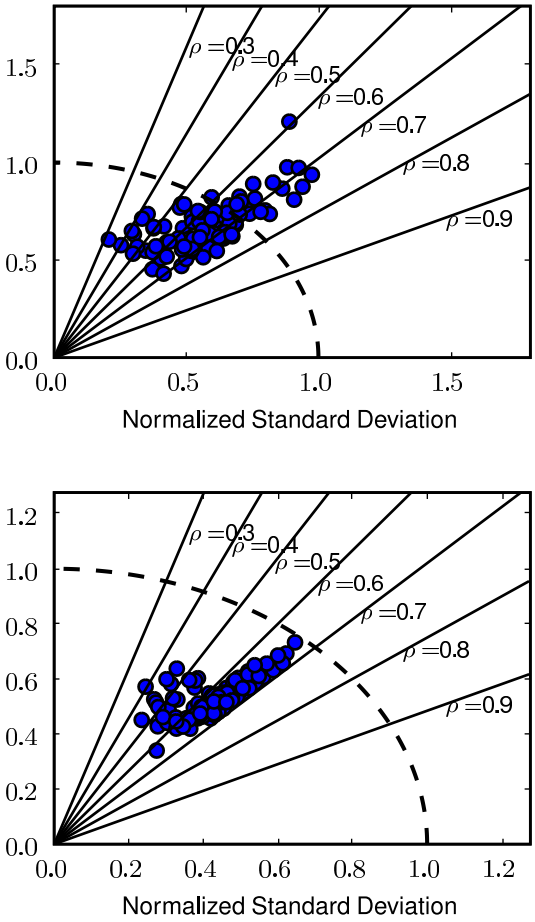


Figure 1: Taylor diagrams for ozone peaks from Airbase stations, for the multimodel ensemble (top) and the Monte Carlo ensemble (bottom).

upper envelope of the ensemble. The rank histogram displays, for each rank, the number of observations with that rank. An ensemble properly estimates the uncertainty if the rank histogram is almost flat.

Figure 2 displays the rank histograms for the multimodel and Monte Carlo ensembles. A significant number of observations fall below the lower envelope of the multimodel ensemble. On the contrary, the Monte Carlo ensemble has a lot of observations above its upper envelope. This shows that the ensembles are not spread enough in given time periods or given regions. Besides the first and last bars, the histograms differ from the flat histogram which would coincide with the dotted lines in the figure.

In order to improve the uncertainty estimation that can be derived from the ensembles, we apply independently to both ensembles the calibration procedure that was introduced in Garaud and Mallet [2011]. The main idea is to extract a sub-ensemble with a flat rank histogram so that its uncertainty estimation should be more accurate. The ensemble calibration is therefore a combinatorial optimization problem. We use a genetic algorithm to select a sub-ensemble that minimizes the variance of the rank histogram [Garaud and Mallet, 2011]. Note that the observations outside the envelope of the full ensemble will be outside the envelope of the sub-ensemble. As a consequence, the heights of the first and last bars in the histogram can only increase after the selection of a sub-ensemble. Since the number of bars (i.e., the number of models plus one) in a flat histogram is the total number of observations divided by the height of one bar, only few members can be part of the sub-ensemble if the bars are too high. In order to increase the number of members in the sub-ensemble, we first remove from each member the global (spatio-temporal) bias of the ensemble mean, so as to decrease the height of the extreme bars. Figure 3 shows the rank histograms of the two unbiased ensembles. The extreme-bars values are significantly lower than with the bias, which eventually helps increasing the number of members selected by the calibration.

Figure 4 displays the rank histograms of the calibrated ensembles. For both ensembles, the rank histograms are almost flat. The calibrated multimodel ensemble includes 34 members while the calibrated Monte Carlo ensemble only has 23 members. This difference can partially be explained by the higher value of the extreme bar in the rank histogram of the full unbiased Monte Carlo ensemble. Also, as was pointed out in section 2.2, the Monte Carlo ensemble shows less variability than the multimodel ensemble, which makes it more difficult to properly span the range of the observations.

2.4 Uncertainty and Covariance Estimation

After calibration, the sub-ensemble is supposed to better sample the distribution of the uncertain ozone concentrations. The empirical standard deviations of the sub-ensembles therefore provide a reliable approximate measure of the uncertainty. If X_i is the state vector of the ith member, in a N-member ensemble, the empirical variance is defined as:

$$\Sigma = \frac{1}{N-1} \sum_{i=1}^{N} \left[X_i - \frac{1}{N} \sum_{j=1}^{N} X_j \right] \left[X_i - \frac{1}{N} \sum_{j=1}^{N} X_j \right]^T . \quad (1)$$

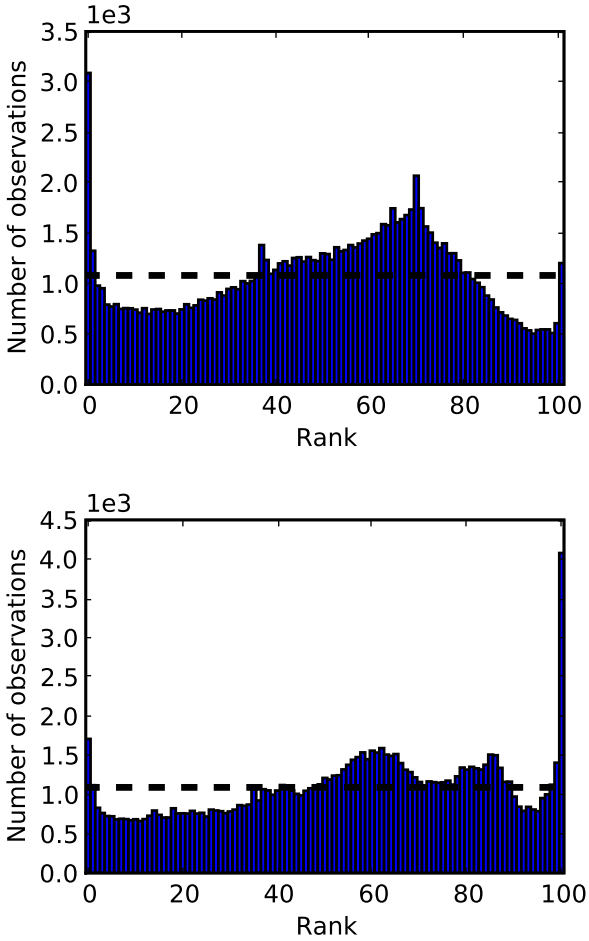


Figure 2: Rank histograms for the multimodel ensemble (top) and the Monte Carlo ensemble (bottom). The dotted line corresponds to the height of the flat rank histogram.

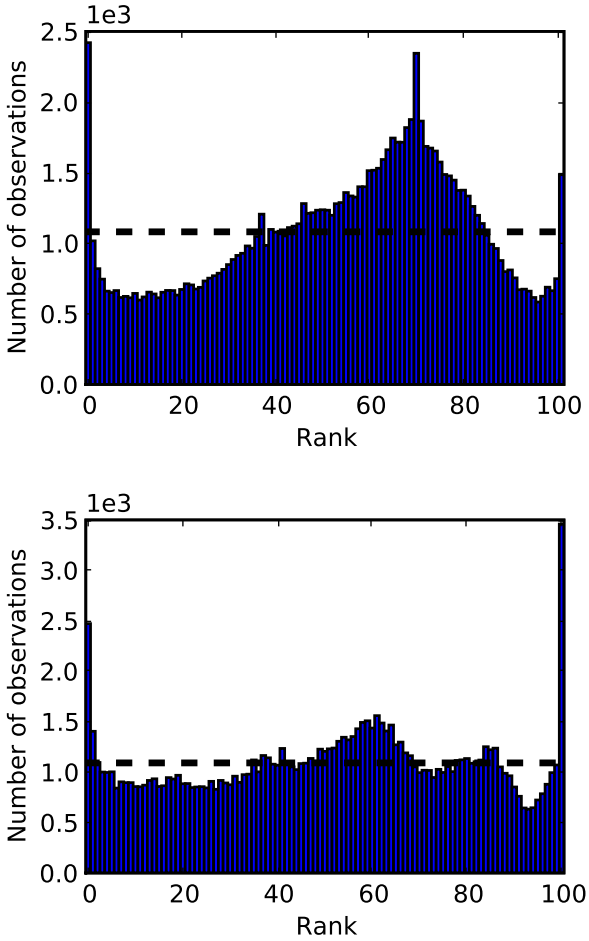


Figure 3: Rank histograms of the multimodel ensemble (top) and the Monte Carlo ensemble (bottom) after removing the bias of the ensemble mean. The dotted line corresponds to the height of the flat rank histogram.

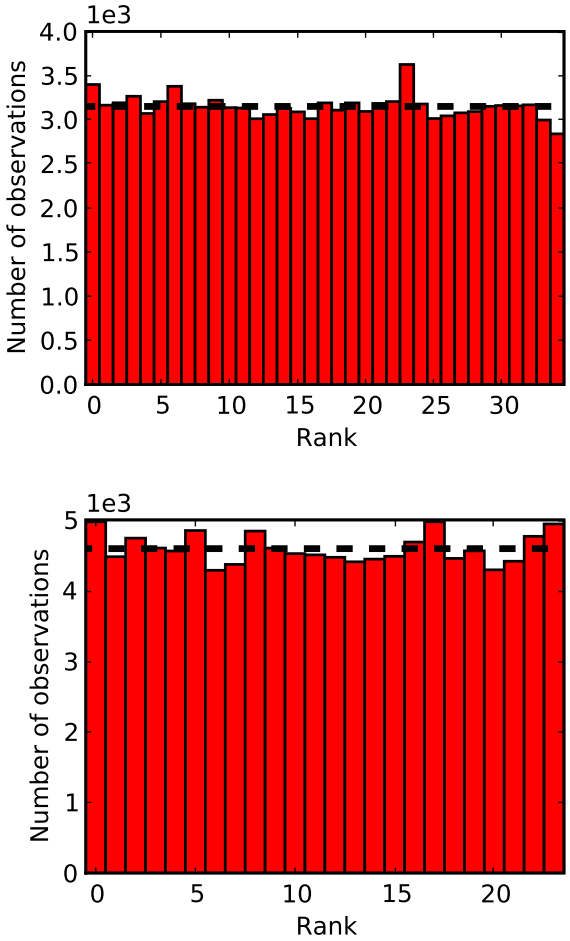


Figure 4: Rank histograms of the multimodel sub-ensemble (top) and the Monte Carlo sub-ensemble (bottom), after calibration. The dotted line corresponds to the height of the flat rank histogram.

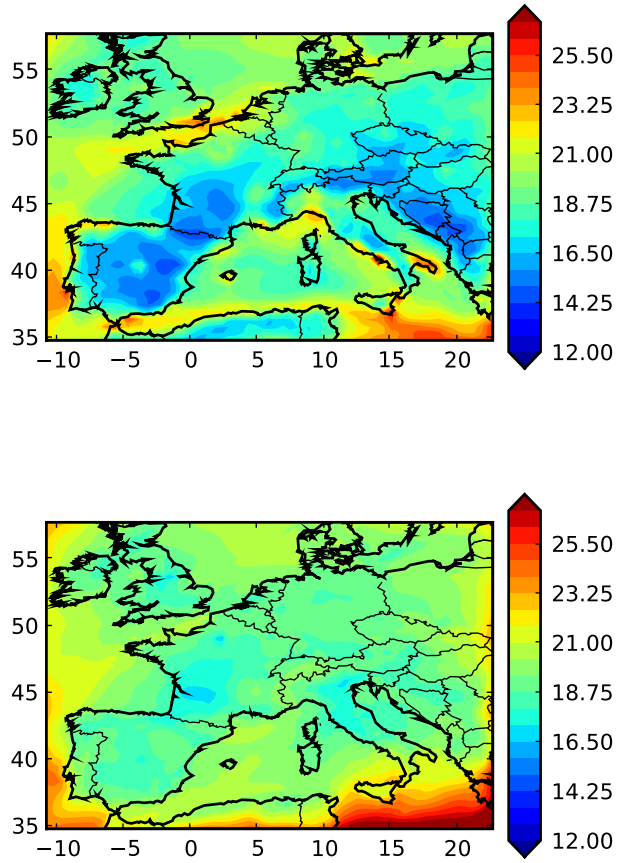


Figure 5: Empirical standard deviation computed from the calibrated multimodel sub-ensemble (top) and calibrated Monte Carlo sub-ensemble (bottom). These fields, in $\mu g\,m^{-3}$, are averaged over May 2001.

Figure 5 shows two maps of the empirical standard deviation (i.e., the diagonal of Σ) averaged over May 2001 for the calibrated multimodel and Monte Carlo sub-ensembles. In the Monte Carlo case, there is a high impact of the uncertain boundary conditions in the south of Europe, and the ozone ensemble standard deviation shows moderate variability in the rest of the domain. In the multimodel case, the field shows more gradients and local high values especially along the costs.

Uncertainty estimation is useful in data assimilation where the corrections on the model's state depend on the amplitudes and on the shapes of the state error variance and the observational error variance. A key point lies in the spatial distribution of the covariance between an error on ozone concentration at a given location and the errors at the other locations. This error covariance corresponds to one line of the matrix Σ. Figures 6 and 7 show two error covariance fields,

computed from the calibrated sub-ensembles. The covariances are computed with respect to two points: one location where there are large emissions (Paris), and another location in the center of France (background).

Monte Carlo covariance fields can show very high values along the domain boundaries. These values can be much higher than the covariances with close locations. It means that the perturbations on the boundary conditions travel down to the inner of the domain, and as the perturbations do not depend on time (nor on space), the transported perturbations remain correlated with those on the boundary conditions. Just like for the variance, the covariances show moderate variability (except along the domain boundary). There are however clear patterns mainly due to the emission points. In figure 6, the covariances with the background cell are lower at emission locations, while in figure 7, the covariance with the point located near emission sources show larger values at all emissions locations (even along the Mediterranean ships route). This raises the so-called localization issue in data assimilation; a data assimilation method that relies on such error covariances would unduly correct the concentrations far away from the observed locations, provided the errors on the model's concentrations at the observed locations are correlated with the errors at these away locations.

The covariances as approximated by the multimodel sub-ensemble show much more variability and the largest values are located in the vicinity of the point with which the covariances are computed. The perturbations in the boundary conditions do not lead to high covariances along the domain boundary, although the relatively high minimum of the covariance (about 200 $\mu g^2\,m^{-6}$) is partially due to the boundary conditions (see section 3.2.2). Finally, as one may expect, the covariance decreases faster (with the distance) for the point located near large emissions than for the background point. As in the Monte Carlo case, most emission locations appear in the maps. This is due to the perturbations on the emissions that do not depend on space and therefore create spurious correlation over long distances.

Compared to the Monte Carlo sub-ensemble, the multimodel sub-ensemble seems to better represent the uncertainty. It includes more models, it is less subject to the covariance localization problem and the patterns in the variance and covariance fields look better (especially along the domain boundaries). In order to improve the results of both approaches, the input data should be perturbed with spatial and temporal decorrelations, instead of with a single space- and time-independent perturbation (multiplicative for log-normal distributions, additive for normal distribution).

3 Uncertainty Due To Input Data

In this section, we evaluate the impact of the uncertainty in the input data on ozone simulated from 1st April to 31th August 2001. We want to identify what are the main uncertainty sources in the temporal mean of ozone peaks at observation stations.

3.1 Correlation and Regression

Following Hanna et al. [2001], a linear regression is carried out between ozone ensemble simulations and the perturbations applied to the input fields and pa-

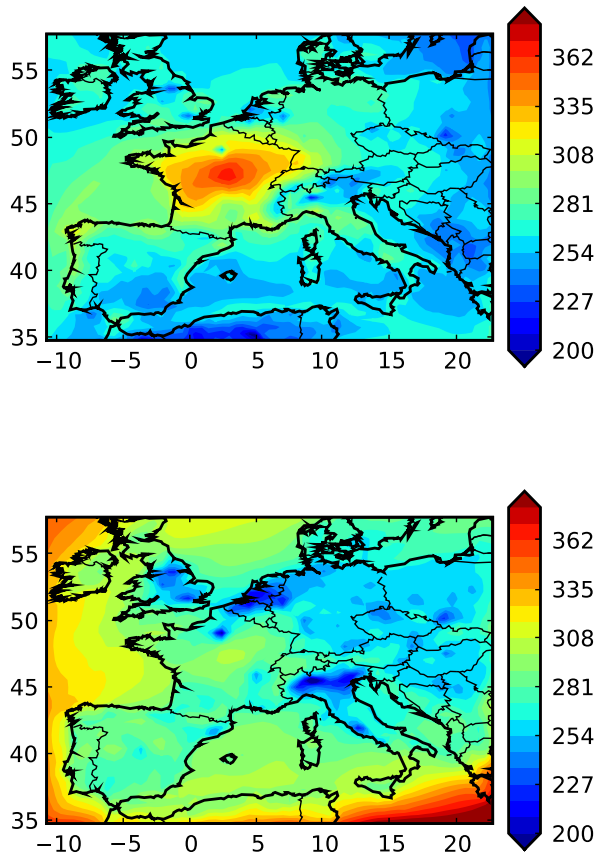


Figure 6: Ozone spatial covariance fields for the multimodel sub-ensemble (top) and the Monte Carlo sub-ensemble (bottom), in $\mu g^2\,m^{-6}$. The location with respect to which the covariance is computed is a background cell in the center of France.

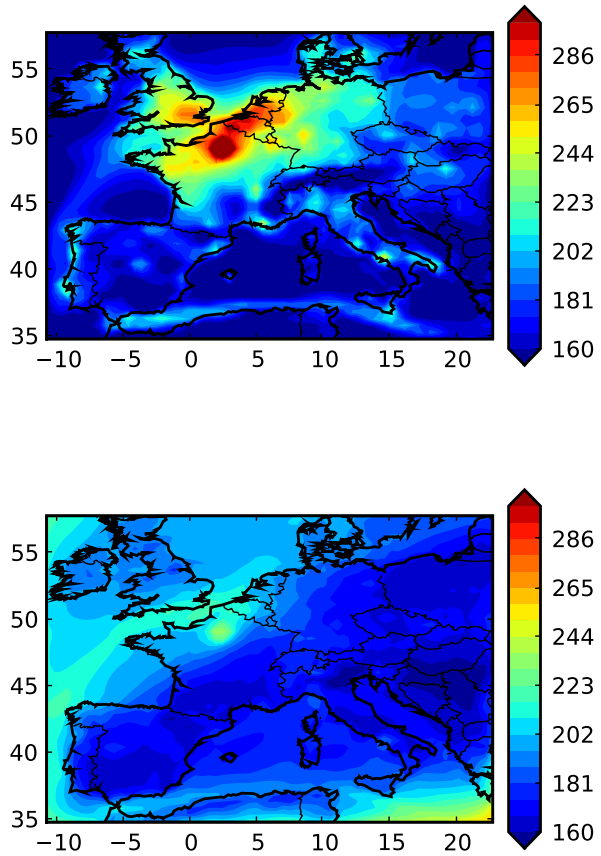


Figure 7: Ozone spatial covariance fields for the multimodel sub-ensemble (top) and the Monte Carlo sub-ensemble (bottom), in $\mu g^2\,m^{-6}$. The location with respect to which the covariance is computed is near Paris (France).

rameters. It identifies the linear relationship between the ozone concentration x_i of simulation i and the perturbations u_{ik} to the input data:

$$x_i = \beta_0 + \beta_1 u_{i1} + \beta_2 u_{i2} + \ldots + \beta_k u_{ik} + \ldots + \beta_K u_{iK} + \epsilon_i \ , \tag{2}$$

where β_k is a regression coefficient to be determined and ϵ_i is the error (i.e., the part of x_i that cannot be explained by the linear regression). u_{ik} is (1) the additive random perturbation ($\frac{1}{2}\alpha\gamma$) applied to the kth uncertain input field if this field is normally distributed, (2) the multiplicative random perturbation ($\sqrt{\alpha}^{\gamma}$) applied to the kth uncertain input field if the field is log-normally distributed, or (3) the multiplicative perturbation $1+\frac{1}{2}\alpha\gamma$ for temperature. ϵ is supposed to be a vector of white noise with a variance $\mathrm{var}(\epsilon) = \sigma_\epsilon^2 I$ where I is the identity matrix.

If $x = [x_1, \ldots, x_i, \ldots, x_N]^T$ is the vector of the ozone concentrations (from a N-member ensemble), $\beta = [\beta_0, \ldots, \beta_k, \ldots, \beta_K]^T$ and $U = [1, u_1, \ldots, u_k, \ldots, u_K]$ is the matrix of all perturbations (except the first column that is filled with ones), the regression reads

$$x = U\beta + \epsilon \ . \tag{3}$$

The regression coefficients are determined as $\beta = (U^T U)^{-1} U^T x$. If the absolute value of a regression coefficient $|\beta_k|$ is high, the ozone concentration is sensitive to the uncertainty on the field k. However for a regression coefficient β_k to be reliable, we require that its estimated value is higher than twice its standard deviation σ_{β_k}. The variance of β is given by

$$\Sigma_\beta^2 = \sigma_\epsilon^2 (U^T U)^{-1} \, . \tag{4}$$

$\sigma_{\beta_k}^2$ corresponds to the k^{th} diagonal term of the matrix Σ_β^2. The variance σ_ϵ^2 is computed from the regression residuals. If a coefficient does not satisfy the criterion $|\beta_k| \geq 2\sigma_{\beta_k}$, it is not excluded from the regression, but we ignore it in subsequent analyses.

The Monte Carlo ensemble contains 100 members in which 285 perturbations (i.e., 285 regressors) are applied. In order to decrease the number of regressors, only the regressors k for which the correlation ρ_k with the ozone concentration is greater than a given threshold are selected. The correlation threshold is taken so that the confidence interval on the correlation between the regressor and the ozone concentration does not include zero. The confidence interval depends on the sample size [Fisher, 1921], hence on the number of ensemble members. With normally-distributed perturbations and ozone concentration, it can be shown [Fisher, 1915] that the 95% confidence interval on the correlation ρ_k is

$$\left[\rho_k - \frac{\mathrm{e}^{2s}-1}{\mathrm{e}^{2s}+1}, \rho_k + \frac{\mathrm{e}^{2s}-1}{\mathrm{e}^{2s}+1}\right] \tag{5}$$

where $s = \frac{1.96}{\sqrt{N-3/2}}$. With $N = 100$, the confidence interval does not include zero if ρ_k is higher than 0.195. Note that most perturbations are not log-normally distributed, so that we only apply the criterion as a guideline.

The correlations are computed independently at each observation station. We then select the perturbations with a correlation satisfying $\rho_k > 0.195$ at over 75% of background, rural or urban stations. It selects 20 parameters for background stations, 19 parameters for rural stations and 18 parameters for

urban stations. Among these selected parameters, one finds ozone boundary conditions, NO_2 photolysis rate, the rate of photolysis of O_3 in O^{1D} (and O_2), attenuation, ozone deposition velocity and a few chemical reactions from RACM. A priori important fields are not always selected by this approach, so whatever their correlation with the output concentrations, we systematically include the perturbations of the nine following variables: VOC emissions, NOx emissions, biogenic emissions, NO and NO_2 boundary conditions, vertical diffusion coefficient, temperature, wind module and wind angle.

After the regression, we compute the determination coefficient R^2. It corresponds to the part of the variance that is explained by the linear combination. If we denote $\hat{x}_i = x_i - \epsilon_i$ and $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$, then the determination coefficient reads

$$R^2 = \frac{\sum_{i=1}^{N}(\hat{x}_i - \bar{x})^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}\,. \tag{6}$$

The determination coefficient depends on the number K of regressors and on the number N of samples. In order to compensate, we use the so-called adjusted determination coefficient AR^2:

$$AR^2 = 1 - (1 - R^2)\frac{N-1}{N-p-1}\,. \tag{7}$$

3.2 Results

The regression is applied independently at each observation station. The regressand x_i is the temporal mean of ozone daily peaks at the station. The regressors depend on the station type (urban, rural or background), since the correlation criterion (see above) selects a slightly different list of parameters for each station type. The regression coefficients and adjusted determination coefficients AR^2 reported in this section are averaged over all stations from one of these types: urban (665 stations), rural (263) or background (92) from Airbase network.

3.2.1 Monte Carlo

Table 2 shows the (averaged) regression coefficients and their spread (empirical standard deviation of the regression coefficients), computed for the most important fields and for each station type. Only the average coefficients that are greater than twice their average standard deviation are reported. The average of the adjusted determination coefficients is the same whatever the station type and is equal to 91%.

All regression coefficients shown in table 2 are in $\mu g\,m^{-3}$. The magnitude of the regression coefficients together with the typical variation of the perturbation coefficients provide a measure of the impact of input uncertainties on the output ozone concentration. For instance, the temperature regression coefficient is equal to 150 $\mu g\,m^{-3}$ and the associated random perturbation ranges in $[-1\%, 1\%]$ (see table 1). Thus, the part uncertainty due to the temperature is up to $\pm 1.5\ \mu g\,m^{-3}$. Another example is ozone boundary conditions whose regression coefficient is $\sim 41\ \mu g\,m^{-3}$ for all stations, with a perturbation ranging in $[1/2, 2]$. The impact of the uncertainty due to ozone boundary conditions

Table 2: Regression coefficients (averaged, in $\mu g\,m^{-3}$) for the three different station types. Output data (regressand) is the temporal average of ozone peaks. Only the average coefficients that are greater than twice their average standard deviation are reported. In brackets, we show the spread among the stations of the coefficients, computed as the empirical standard deviation of all regression coefficients – and not as the mean of the deviations σ_{β_k}.

Field Name	**Background**	**Rural**	**Urban**
O_3 bounday conditions	41.4 (4)	41.8 (4.2)	40.3 (4.3)
NO_2 photolysis	15.8 (10.8)	17.5 (8.8)	18.5 (9.6)
NOx emission	10.7 (5.1)	4.1 (7)	6. (8.8)
Vertical diff. coeff.	4. (2.4)	5.6 (1.8)	5. (3)
ISO emission	2.5 (2.9)	3.9 (2.2)	3.7 (3.1)
Attenutation	11.7 (5.3)	13.6 (5.1)	15 (5.9)
Temperature	162.3 (33.5)	147.7 (32)	160.5 (34)
API boundary conditions	5. (1.7)	–	–
NO_2 boundary conditions	1.8 (0.4)	1.8 (0.4)	1.6 (0.6)
Wind module	–	–	-6.3 (5.7)
Reaction 28	-32.1 (7.6)	-32.3 (8.3)	-32 (10)
Reaction 90	32.8 (9)	26.4 (9.8)	25.7 (11.2)
Reaction 182	-30 (6)	–	–

therefore ranges roughly in $[-20.5, +41.0]\mu g\,m^{-3}$. Table 3 shows this impact for all input fields which appear in table 2.

For any field except NOx emissions, the regression coefficients are essentially the same for all station types. NOx emissions are given a higher coefficient for background stations than rural or urban stations. NOx are the limiting species for ozone production in background areas. Hence the sensitivity of ozone concentrations to NOx emissions is higher in these areas.

Ozone boundary conditions show an especially high positive regression coefficients, which is consistent with the high uncertainties along the domain boundaries that were pointed out in section 2.4.

The RACM chemical reactions 28 and 90 (see R 1 and R 2) have a significant regression coefficient for all station types. The former has a negative impact with a coefficient at $-32\ \mu g\,m^{-3}$, i.e., an impact on ozone ranging in $[-3.0, +2.9]\mu g\,m^{-3}$. On the contrary, chemical reaction 90 has a positive impact on ozone.

$$O^{1D} + H_2O \longrightarrow HO + HO \qquad (R\,1)$$

$$ONIT + HO \longrightarrow HC_3P + NO_2 \qquad (R\,2)$$

Chemical reaction 28 has a negative impact, which is difficult to explain since increased concentrations of HO should lead to more ozone production. The latter is confirmed by the sensitivities computed by Martien and Harley [2006] in the context of 3D photochemical simulations.

Table 3: Regression coefficients and the associated 95% uncertainty range on ozone ($\mu g\, m^{-3}$) for several input fields described in table 1. The regression coefficients are averaged over all stations whatever their type, except for NOx emissions and reaction 182 whose regression coefficients are averaged at background stations only. The uncertainty range represents the deviation of the output concentration with respect to its median.

Field Name	**Averaged Coeff.**	**Uncertainty Range**
O_3 boundary conditions	41.0	-20.5, +41.0
NO_2 photolysis	17.0	-4.9, +6.8
NOx emissions	10.7	-4.4, +7.5
Vertical diff. coeff.	5.0	-2.4, +4.5
ISO emissions	3.0	-1.5, +3.0
Attenutation	13.0	-3.0, +3.9
Temperature	150.0	-1.5, +1.5
API bounday conditions	5.0	-2.5, +5.0
NO_2 boundary conditions	1.8	-1.2, +3.6
Wind module	-6.0	-3.0, +2.0
Reaction no. 28	-32.0	-3.2, +2.9
Reaction no. 90	30.0	-2.7, +3.0
Reaction no. 182	-30.0	-3.0, +2.7

It is easier to explain why reaction 90 (R 2) has a positive regression coefficient, since it tends to indirectly produce ozone through the production of NO_2.

RACM chemical reaction 182 (R 3) has a negative regression coefficient which is significant only for background station type.

$$\mathrm{TOLP} + \mathrm{MO_2} \longrightarrow \mathrm{HCHO} + \mathrm{HO_2} + 0 \cdot 35\,\mathrm{MGLY} + 0 \cdot 65\,\mathrm{GLY} + \mathrm{DCB} \quad \text{(R 3)}$$

Species TOLP tends to produce ozone via RACM chemical reactions 41, 142 or 202 [Stockwell et al., 1997]. Thus, if the reaction rate of R 3 increases, TOLP concentration decreases and indirectly tends to produce less ozone.

Note that, even if we can explain the sign of the regression coefficients associated to the previous chemical reactions, it is however difficult to explain why these particular reactions among many others lead to higher uncertainties in ozone concentrations.

3.2.2 Multimodel Ensemble

In the multimodel case, all perturbations are included in the regression, since fewer input fields are perturbed. In particular, the rates of the chemical reactions are not perturbed. The perturbed fields are:

- VOC boundary conditions and VOC emissions;
- NO_x boundary conditions and NO_x emissions;
- O_3 boundary conditions;

Table 4: Regression coefficients (averaged, in $\mu g\,m^{-3}$) for the three different station types. Output data (regressand) is the temporal average of ozone peaks from multimodel ensemble. Only the average coefficients that are greater than twice their average standard deviation are reported. In brackets, we show the spread among the stations of the coefficients, computed as the empirical standard deviation of all regression coefficients – and not as the mean of the deviations σ_{β_k}.

Field Name	**Background**	**Rural**	**Urban**
O3 Boundary Conditions	42 (4.6)	45 (5)	43.8 (5)
NOx Emissions	21 (7)	13.2 (9.8)	15.3 (10.9)
Photolysis	17.9 (5.8)	17.8 (4.8)	18.8 (6.5)
Biogenic emissions	7.8 (3.)	7.8 (2.2)	8.3 (3.2)
Wind module	-14 (3.5)	-14.4 (5.5)	-14 (6)
Deposition	-14 (4.4)	-12.8 (3.6)	-11.8 (4.2)
NOx Boundary Conditions	3 (0.8)	–	2.7 (0.9)

- biogenic emissions.
- vertical diffusion, temperature and wind;
- deposition velocities;
- photolysis rates.

Note that the same perturbation is applied to all deposition velocities, whatever the chemical species. It was assumed that these velocities where highly correlated since they are based on the same aerodynamic resistance, quasilaminar sublayer resistance and since they are based on a bulk surface resistance derived from those of SO_2 and O_3 [Wesely, 1989; Zhang et al., 2003]. The perturbations of the photolysis rates are also the same for all species. The perturbation scheme is described in Garaud and Mallet [2010].

Table 4 shows the eight significant averaged regression coefficients. Ozone boundary conditions have the highest regression coefficient, as in Monte Carlo case. It is noteworthy that ozone boundary conditions and photolysis rates are associated with similar regression coefficients as with the Monte Carlo ensemble: $\sim 44\ \mu g\,m^{-3}$ and $\sim 18\ \mu g\,m^{-3}$ respectively. The adjusted determination coefficient is equal to 72% — instead of 91% for the Monte Carlo case. This is consistent with the fact that part of uncertainty in the multimodel ensemble is represented by changes in the model formulation itself.

The average regression coefficient for deposition velocities is negative. When all deposition velocities increase, ozone deposition should play a key role and ozone concentrations should therefore decrease. The average regression coefficient for wind velocity is also negative, which is probably due to emitted pollutants (especially ozone precursors) being more diluted when the wind velocity increases.

As in the Monte Carlo study, the average regression coefficient for NOx emissions is higher at background stations than at urban and rural stations.

Whatever the ensemble, the part of the uncertainty due to ozone boundary conditions is high. This is consistent with the previous uncertainty maps, figures 5 and 6. It indicates that a multimodel ensemble, or at least a Monte Carlo ensemble, should be beneficial at higher scale, in order to provide an ensemble of boundary conditions. The description of the uncertainty in the boundary conditions would then be much more accurate than the perturbation scheme we use in this work.

Now that we have analyzed the main sources of uncertainties due to the input fields and parameters, we investigate in the next section how the discrepancies between the observations and the simulations can be decomposed and what part is due to the shortcomings of the modeling.

4 Error Decomposition

In this section, we investigate the observational errors, we try to evaluate the representativeness error, and we derive from them the modeling error. Consider a model' state vector X and a vector of observations Y at the same time. Both can be compared using the observation operator that maps the state space into observation space, so that HX can be compared to Y. If the true concentration vector at the observed locations is Y^t and the true state vector is X^t, the discrepancy between the observation vector and its simulated counterpart can be decomposed as follows

$$\begin{aligned} e &= Y - HX \\ &= \underbrace{Y - Y^t}_{\text{measurement}} + \underbrace{Y^t - HX^t}_{\text{representativeness}} + H\underbrace{(X^t - X)}_{\text{modeling}} \\ &= e_o + e_r + He_m \,. \end{aligned} \tag{8}$$

The measurement error $e_o = Y - Y^t$ is notably due to limitations of the observation instrument, errors in the calibration of the instrument, possible mistakes in the retrieval and errors in the postprocessing of the raw measurements. The modeling error $e_m = X^t - X$ is due to the shortcomings of the model that computed the state X based on approximate physical, chemical and numerical formulation and erroneous input data. Even if X^t is known, it is not possible to compute the exact concentrations at the observation locations. The exact state vector X^t contains concentrations averaged in the model's grid cells, from which one cannot compute a point concentration inside a grid cell because of the sub-grid variability. We refer to a representativeness error $e_r = Y^t - HX^t$. Note that the relation between the state vector and the observations is provided by H which can be another source of errors. In our case, H is simply a linear operator (the 2D concentration field X being bilinearly interpolated at observed locations), but in general, H can be a complex operator based on approximations.

Our objective is to estimate the variance of each error. First, we assume that the errors e_o, e_r and e_m have zero mean and that they are mutually uncorrelated, e.g., $\mathrm{E}[H(X^t - X)(Y - Y^t)^T] = 0$. If H_i is the ith row of H, the covariance

between two error components i and j of e is

$$\begin{aligned}
\mathrm{Cov}(e)_{ij} = \mathrm{E}([&(Y_i - Y_i^t) + (Y_i^t - H_i X^t) + (H_i X^t - H_i X)] \\
&[(Y_j - Y_j^t) + (Y_j^t - H_j X^t) + (H_j X^t - H_j X)]^T) \\
= \quad & \underbrace{\mathrm{E}[(Y_i - Y_i^t)(Y_j - Y_j^t)^T]}_{\text{measurement error variance}} \\
+ \quad & \underbrace{\mathrm{E}[(Y_i^t - H_i X^t)(Y_j^t - H_j X^t)^T]}_{\text{representativeness error variance}} \\
& + H_i \underbrace{\mathrm{E}[(X^t - X)(X^t - X)^T]}_{\text{modeling error variance}} H_j^T
\end{aligned} \tag{9}$$

4.1 Measure Error

In this study, we use ground stations from the European Airbase network. Ozone is measured by spectrometry, using its absorption in ultraviolet. A sample of ambient air is taken. A beam at wavelength 254 nm is emitted. Ozone molecules absorb a part of the radiation. A sensor turns the measured radiation into an electrical signal which is proportional with the sampling ozone concentration. Airparif, the organization responsible for monitoring air quality in the Paris region, produces upper bounds on the measurement uncertainties, along with the measurements themselves [Airparif, 2007]. The uncertainty takes into account many error sources from the different stages of the measurement chain: air sampling, data capture, electronic device, calibration, ...

Hourly ozone measurements and their uncertainty upper bounds, provided by Airparif for year 2009 and 30 monitoring stations, are clustered in a concentration intervals from $[0, 20]\mu\mathrm{g\,m^{-3}}$ to $[80, \infty[\mu\mathrm{g\,m^{-3}}$. The uncertainty and the relative uncertainty (i.e., the uncertainty divided by the concentration) are reported in table 5 and in figure 8. The uncertainty corresponds to the 95% confidence interval, so that, in case of Gaussian errors, it is equal to twice the standard deviation. The uncertainty increases with the concentration, while the relative uncertainty decreases. The relative uncertainty can be higher than 50% when the measured ozone concentration is about 8 $\mu\mathrm{g\,m^{-3}}$. For high concentrations, the relative uncertainty ranges from $\sim$ 12% to $\sim$ 9% for ozone concentrations between ~ 80 and ~ 150 $\mu\mathrm{g\,m^{-3}}$.

4.2 Modeling and Representativeness Errors

In this section, we carry out two independent methods to estimate the variance of the representativeness error. We consider all ozone hourly observations (not only the ozone peaks) and, in the second method, the multimodel-ensemble mean.

The first method is solely based on the observations. We assume that the mean concentration in a grid cell can be approximated by the mean of the observed concentrations. This assumption is reasonable only if there are enough observation stations in one model grid cell, and if they are spread all over the grid cell. In grid cell k, we denote J_k the set of the indexes of stations that are inside the grid cell. We approximate the mean concentration in the grid cell by

Table 5: For five concentration intervals, the table shows the average ($\mu g\,m^{-3}$) of all measurements in the interval, the corresponding average of the uncertainty ($\mu g\,m^{-3}$) and the corresponding relative uncertainty. The uncertainty corresponds to a 95% confidence interval; so if the error is Gaussian, it is twice the standard deviation.

Range ($\mu g\,m^{-3}$)	**Av. Measure**	**Uncertainty**	**Rel. Unc.**
0 – 20	8.8	6.8	0.77
20 – 40	30.4	7.4	0.24
40 – 60	49.7	8.1	0.16
60 – 80	68.5	8.9	0.13
≥ 80	97.9	10.4	0.11

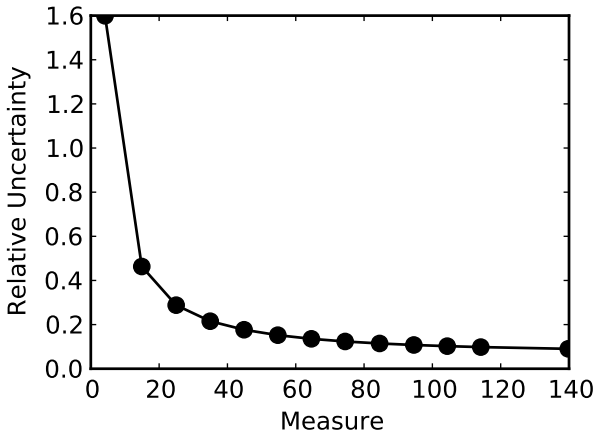


Figure 8: Relative uncertainty according to the observed ozone concentration in $\mu g\,m^{-3}$.

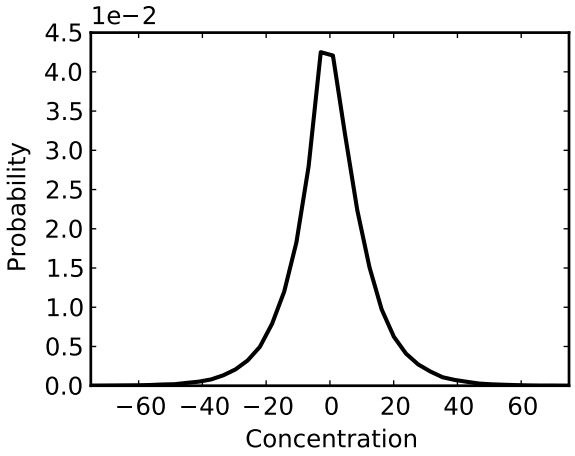


Figure 9: Approximate relative frequency occurrence of the representativeness error. The empirical standard deviation is 12.9 $\mu g\,m^{-3}$.

$$A_k = \frac{1}{|J_k|} \sum_{j \in J_k} Y_j \ , \tag{10}$$

where $|J_k|$ is the number of stations inside the grid cell k.

We select six grid cells that contain between 8 and 10 Airbase stations: two cells close to Marseille, one close to Paris, Barcelona, Valencia and London. All selected stations are urban, since in rural regions, the monitoring network is not dense enough to have so many stations in one grid cell.

We first compute $A_k - Y_j$ for the eight grid cells of interest and for all corresponding station $j \in J_k$, which amounts to $300,000$ discrepancies. This quantity measures how much observations can deviate from the approximate average in the cell. Figure 9 shows the relative occurrence frequency of $A_k - Y_j$. By definition of A_k, the mean of the distribution is zero. The empirical standard deviation, which provides an estimate of the standard deviation of the representativeness error, is equal to 12.9 $\mu g\,m^{-3}$.

The second method is the sometimes referred to as the observational method or the Hollingsworth-Lönnberg method [Hollingsworth and Lönnberg, 1986]. The variance of the modeling error and the sum of measurement and representativeness variances are estimated based on a variogram of the discrepancies $Y - HX$. The variogram plots the empirical covariance between all pairs $(Y_i - H_i X, Y_j - H_j X)$ against the distance between the locations i and j. The first bar of the diagram, which corresponds to variances (because the distance is zero), is due to all three errors e_o, e_r and e_m. If the observational errors are assumed to be uncorrelated, the height of the next bars is only due to the modeling error e_m. If one extrapolates from these bars to the origin, the difference between the ordinate at the origin and the height of the first bar is due to the observational error $(e_o + e_r)$ — see figure 10 for an illustration.

In our case, we consider all pairs (i, j) of urban stations, and all times at which observations are available. The error is computed using the mean of the

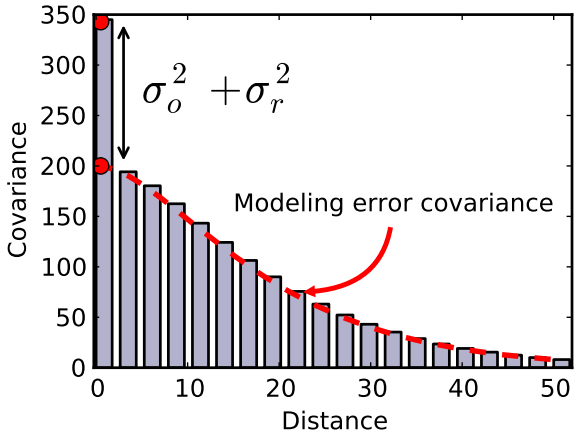


Figure 10: Illustration of the method by Hollingsworth and Lönnberg [1986] to estimate observational error variance with a variogram. This figure is inspired by Bouttier and Courtier [1999].

calibrated multimodel sub-ensemble.

In total, $\sim 190,000$ covariances are computed. Figure 11 shows all covariances that decrease with the distance as the errors become uncorrelated.

We collect all covariance values from points within a distance in $]0, 0.5°]$. The mean of these covariances is equal to $\sigma_m^2 \simeq 336\ \mu g^2\,m^{-6}$. The mean of the variance (computed with all pairs with null distance) is $\sigma^2 = 489\ \mu g^2\,m^{-6}$. Hence $\sigma_o^2 + \sigma_r^2 \simeq \sigma^2 - \sigma_m^2 \simeq 153\ \mu g^2\,m^{-6}$. The mean of observed concentrations is about $43\ \mu g\,m^{-3}$. According to section 4.1, it means that the variance of measurement errors is about $16\ \mu g^2\,m^{-6}$. Consequently, the variance of the representativeness error can be estimated by $137\ \mu g^2\,m^{-6}$, hence a standard deviation at $11.7\ \mu g\,m^{-3}$ which is a bit less than the estimation from the first method ($12.9\ \mu g\,m^{-3}$).

Observational error should be independent of the model, provided the same observation operator is used, which is the case in our ensemble where all models have the same horizontal resolution. We carry out the method with seven models randomly selected from the multimodel ensemble. The estimated variance of the observational error varies between $146\ \mu g^2\,m^{-6}$ and $156\ \mu g^2\,m^{-6}$. The standard deviation of the representativeness error is then estimated between 11.7 and $12.5\ \mu g\,m^{-3}$.

Note that the measurement and representativeness errors are not entirely uncorrelated between two points at close distance. For instance, the measurement errors can depend on the atmospheric conditions, which are obviously correlated at short distance. Therefore the estimation of the modeling error variance is overestimated, and the representativeness errors is underestimated. According to this method, the standard deviation of the representativeness error is likely to be greater than $12.5\ \mu g\,m^{-3}$, which is consistent with the first method giving $12.9\ \mu g\,m^{-3}$.

The variance of the discrepancies between hourly ground-level ozone observations and the mean of the calibrated sub-ensemble is $489\ \mu g^2\,m^{-6}$. Follow-

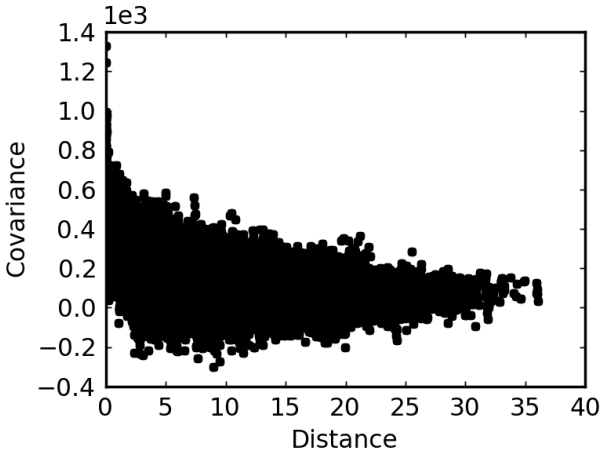


Figure 11: Variogram of the error between observed and simulated ozone concentration. The covariance is in $\mu g^2\,m^{-6}$ and the distance in degrees (latitude/longitude).

ing (9), it can be decomposed in less than $\sim 3.2\%$ for measurement errors ($\sigma_o^2 \lesssim 16\ \mu g^2\,m^{-6}$), in about 34% for representativeness error ($\sigma_r^2 \simeq 166\ \mu g^2\,m^{-6}$) and in about 63% for modeling error ($\sigma_m^2 \simeq 489 - 16 - 166 = 307\ \mu g^2\,m^{-6}$).

5 Discussion and Conclusions

Two methods for the generation of ensembles are compared in this paper: (1) the Monte Carlo method where several input fields and parameters are perturbed and (2) a multimodel ensemble generation which takes into account uncertainties in input data, numerical approximations and chemical/physical parameterizations. Monte Carlo simulations show less variety than the simulations from the multimodel ensemble. A posteriori calibrations are carried out on both ensembles in order to get more accurate uncertainty estimations. Even after the calibration, the multimodel sub-ensemble shows better features than the Monte Carlo sub-ensemble. More members are selected, and the variances and covariances patterns look better (especially along the domain boundaries). These results seem to justify the use of a multimodel ensemble in place of Monte Carlo simulation whenever it is possible. Note that in this paper, the multimodel ensemble also includes perturbations in the input data and it has a large number of members. A multimodel ensemble with just a few members might not show the same favorable features since the estimation of a variance may require more than a dozen members. It would however be interesting to extend this study with a multimodel ensemble based on models from different teams.

The regressions show that the uncertainties on the ozone boundary conditions, among all other input fields and parameters, have the largest effect on the uncertainties on output ozone concentrations. However, many other parameters and fields have a significant impact on ozone concentrations, which

leads to a rather high total uncertainty. The mean uncertainty on ozone hourly concentrations, computed as twice the mean standard deviation divided by the mean concentration, is as high as 57.5%, according to the calibrated multimodel sub-ensemble over the full year 2001.

In the last section, we split the discrepancy between observations and simulations into a measurement error, a representativeness error and a modeling error. The measurement error is comparatively low. The representativeness error appears to be rather high, since it explains about 34% (in terms of variance) of the discrepancy with the observations. Consequently, the errors in the model outputs are significantly lower than the raw comparison with the observations suggests.

In future work, the generation of the ensembles could be improved. For all input fields, the perturbations should depend on time and space, with decorrelation length and time to be determined. The multimodel ensemble should be based on even more alternatives in the models generation. It is important that modern modeling systems get more flexibility so that they can include a wider range of model formulations. This especially applies to aerosol simulations, which were not addressed in this paper. There should be an extended range of options in the multimodel ensemble since many uncertainties lie in the formulation of the aerosol dynamics.

Finally, it would be interesting to evaluate if the uncertainty estimations computed in this study could help improving the performance of data assimilation. For instance, optimal interpolation could benefit from the patterns of the empirical state error variances. Another example is the ensemble Kalman filter that could rely on a multimodel ensemble instead of the usual Monte Carlo ensemble.

Acknowledgement

We would like to thank to Hélène Marfaing and Christophe Debert from Airparif for their very useful studies and their data about measurement uncertainties. We thank Christian Seigneur for his insights on the chemical processes.

References

Airparif (2007). Guide pratique d'utilisation pour l'estimation de l'incertitude de mesure des concentrations en polluants dans l'air ambiant. Technical Report Version 9, AIRPARIF.

Anderson, J. L. (1996). A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *Journal of Climate*, 9(7):1518–1530.

Beekmann, M. and Derognat, C. (2003). Monte Carlo uncertainty analysis of a regional-scale transport chemistry model constrained by measurements from the atmospheric pollution over the Paris area (ESQUIF) campaign. *Journal of Geophysical Research*, 108(D17):8,559.

Bouttier, F. and Courtier, P. (1999). Data assimilation concepts and methods. Meteorological training course lecture series, ECMWF.

Boynard, A., Beekmann, M., Forêt, G., Ung, A., Szopa, S., Schmechtig, C., and Coman, A. (2011). An ensemble assessment of regional ozone model uncertainty with an explicit error representation. *Atmospheric Environment*, 45(3):784–793.

Delle Monache, L. and Stull, R. B. (2003). An ensemble air-quality forecast over western Europe during an ozone episode. *Atmospheric Environment*, 37:3,469–3,474.

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, 10(4):507–521.

Fisher, R. A. (1921). On the "probable error" of a coefficient of a correlation deduced from a small sample. *Metron*, 1(4):3–32.

Garaud, D. and Mallet, V. (2010). Automatic generation of large ensembles for air quality forecasting using the Polyphemus system. *Geoscientific Model Development*, 3(1):69–85.

Garaud, D. and Mallet, V. (2011). Automatic calibration of an ensemble for uncertainty estimation and probabilistic forecast: Application to air quality. *Journal of Geophysical Research*, 116(D19304).

Hamill, T. M. and Colucci, S. J. (1997). Verification of Eta/RSM short-range ensemble forecasts. *Monthly Weather Review*, 125:1312–1327.

Hanna, S. R., Chang, J. C., and Fernau, M. E. (1998). Monte Carlo estimates of uncertainties in predictions by a photochemical grid model (UAM-IV) due to uncertainties in input variables. *Atmospheric Environment*, 32(21):3,619–3,628.

Hanna, S. R., Lu, Z., Frey, H. C., Wheeler, N., Vukovich, J., Arunachalam, S., Fernau, M., and Hansen, D. A. (2001). Uncertainties in predicted ozone concentrations due to input uncertainties for the UAM-V photochemical grid model applied to the July 1995 OTAG domain. *Atmospheric Environment*, 35(5):891–903.

Hogrefe, C., Rao, S. T., Kasibhatla, P., Hao, W., Sistla, G., Mathur, R., and McHenry, J. (2001). Evaluating the performance of regional-scale photochemical modeling systems: Part II – ozone predictions. *Atmospheric Environment*, 35:4,159–4,174.

Hollingsworth, A. and Lönnberg, P. (1986). The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: the wind field. *Tellus*, 38A:111–136.

Lattuati, M. (1997). *Contribution à l'étude du bilan de l'ozone troposphérique à l'interface de l'Europe et de l'Atlantique Nord: Modélisation lagrangienne et mesures en altidude*. PhD thesis, Université Paris 6.

Mallet, V., Quélo, D., Sportisse, B., Ahmed de Biasi, M., Debry, É., Korsakissok, I., Wu, L., Roustan, Y., Sartelet, K., Tombette, M., and Foudhil, H. (2007). Technical Note: The air quality modeling system Polyphemus. *Atmospheric Chemistry and Physics*, 7(20):5,479–5,487.

Mallet, V. and Sportisse, B. (2006). Ensemble-based air quality forecasts: A multimodel approach applied to ozone. *Journal of Geophysical Research*, 111(D18).

Martien, P. T. and Harley, R. A. (2006). Adjoint sensitivity analysis for a three-dimensional photochemical model: application to Southern California. *Environmental Science & Technology*, 40(13):4,200–4,210.

McKeen, S., Chung, S. H., Wilczak, J., Grell, G., Djalalova, I., Peckham, S., Gong, W., Bouchet, V., Moffet, R., Tang, Y., Carmichael, G. R., Mathur, R., and Yu, S. (2007). Evaluation of several $PM_{2.5}$ forecast models using data collected during the ICARTT/NEAQS 2004 field study. *Journal of Geophysical Research*, 112(D10S20).

Pinder, R. W., Gilliam, R. C., Appel, K. W., Napelenok, S. L., Foley, K. M., and Gilliland, A. B. (2009). Efficient probabilistic estimates of surface ozone concentration using an ensemble of model configurations and direct sensitivity calculations. *Environmental Science & Technology*, 43(7):2,388–2,393.

Russell, A. and Dennis, R. (2000). NARSTO critical review of photochemical models and modeling. *Atmospheric Environment*, 34:2,283–2,234.

Stockwell, W. R., Kirchner, F., Kuhn, M., and Seefeld, S. (1997). A new mechanism for regional atmospheric chemistry modeling. *Journal of Geophysical Research*, 102(D22):25,847–25,879.

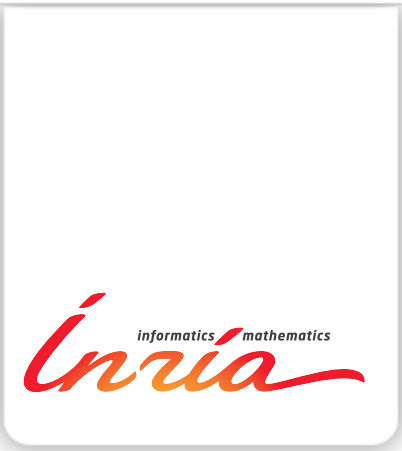
Talagrand, O., Vautard, R., and Strauss, B. (1999). Evaluation of probabilistic prediction system. Proceedings of the ECMWF Workshop on Predictability.

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research*, 106(D7):7,183–7,192.

US EPA (1991). Guideline for regulatory application of the urban airshed model. Technical Report EPA-450/4-91-013, US EPA.

Wesely, M. L. (1989). Parameterization of surface resistances to gaseous dry deposition in regional-scale numerical models. *Atmospheric Environment*, 23:1,293–1,304.

Zhang, L., Brook, J. R., and Vet, R. (2003). A revised parameterization for gaseous dry deposition in air-quality models. *Atmospheric Chemistry and Physics*, 3:2,067–2,082.



RESEARCH CENTRE
PARIS – ROCQUENCOURT

Domaine de Voluceau, - Rocquencourt
B.P. 105 - 78153 Le Chesnay Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399